Journal of Research High School

Vol. 2023 (2) 40 – 47
ISSN 2688-3651 [online]

JRHS Outstanding Research Paper Award

An AI-Powered Assistive Device for the Visually Impaired

Katherine Hua[1] *

[1]Woodbridge High School, Irvine, CA, USA
*Corresponding Author: shivam.mohanty08@gmail.com

Advisor: Andrew Gibas, AndrewGibas@iusd.org

Received April 13, 2022; Revised June 8, 2023; Accepted, June 28, 2023

Abstract

Global studies report that 253 million people suffer from visual impairment. Most rely on traditional aids including Braille, white canes and guide dogs which lack versatility and adaptability. This study intended to utilize A.I. text, object recognition models and ultrasonic technology to create an effective vision aid. The prototype was constructed using a Raspberry Pi board with a Pi camera, ultrasonic sensor, earbuds, and other peripherals. Text and object recognition algorithms were implemented to convert printed text and real objects captured by the camera's live video feed to text output. Then, the text-to-speech code programmed in the device helped convert its text output to speech that could be heard through earbuds. Additionally, the ultrasonic sensors were programmed to determine the distance to the objects by measuring the time between emitting and receiving the reflected ultrasonic waves. During testing, the system recognized all sample words in 2.5 seconds on average and sample sentences in 6.4 seconds with 100% accuracy. Additionally, the device took 2.9 seconds on average to detect 80% of tested objects accurately and could detect large objects such as cars as far as 584cm away. The functional testing indicated that the prototype could inform users of recognized words and sentences, the types of detected objects, and their distance through audio. Thus, the results support the hypothesis that an AI-powered electronic device has the potential to provide reliable visual assistance to the visually impaired in their daily life.

Keywords: Visually impaired, Raspberry Pi, Vision aid, OCR, Text recognition, Object recognition, Ultrasonic sensor

1. Introduction

Vision plays a significant role in our daily lives. Our eyes allow us to explore and understand the physical world, learn new things, avoid danger and hazardous objects, and interact with other people. According to the Vision Loss Expert Group 2015 study, thirty-six million people suffer from blindness and an additional 217 million have medium to severe visual impairment (Bourne, et al., 2017). A report conducted by the National Center for Health Research found that one in five visually impaired people was unable to perform personal care activities while 46% had limited ability (National Center for Health Research, 2004). Their vision impairment causes them physical suffering, loss of productivity, lack of self-confidence, and an inferior quality of life. The economic burden of vision loss in the United States was estimated to be $134.2 billion in 2017, which included $98.7 billion in direct costs such as medical expenses for an eye examination, diagnosis, and corrections, and indirect costs such as informal care and social assistance programs (Rein et al., 2022). There is both a significant humanitarian demand and financial motive for developing a visual assistive aid using current-gen technology to grant much greater in-depth navigational and interpretive capability for the visually impaired. The question is, can it be achieved?

The visually impaired currently are provided with traditional aids such as Braille, white canes, and guide dogs as assistance in their daily life. However, these traditional mechanisms have major disadvantages and flaws. Since Braille is designed to function solely through touch and tactile recognition, it can take some time to develop the touch sensitivity, and practice using braille. This may become a challenge for seniors with slower cognitive and mobile abilities**.** While white canes help users locate obstacles to avoid, they cannot detect mobile objects or out-of-reach

Vol. 2023 (2) 40 – 47
ISSN 2688-3651 [online]

hazards. Guide dogs are more versatile and responsive to situations, but they cost $32,000 for training and annual maintenance and can only work for 6-8 years (*The cost of a service dog*, 2022). Thus, a smart tool that has longevity and employs modern assistive technology to help the visually impaired is urgently needed.

Computer vision is one of the most widely used Artificial Intelligence applications and allows computers to gain a high-level understanding of contents from digital images and videos. The two core pre-existing technologies behind computer vision are deep learning and convolutional neural networks, known as CNN. CNNs are the core of computer vision as they are specifically modeled like the human brains to be able to recognize patterns and interpret visual data. Meanwhile, deep learning is a type of machine learning with algorithms structured in a hierarchy of increasing complexity and abstraction to develop high-level comprehension. It attempts to mimic the human mind's capabilities and thought processes through these trained algorithms (Voulodimos, et al., 2018).

Text and object recognitions are the two most prominent applications of AI in assistive aid for the visually challenged. While humans can quickly identify letters and words based on their ingrained knowledge and their trained visual recognition, machines do not have these natural abilities. Instead, computer vision allows machines to utilize image processing algorithms to understand the text within images through OCR, or Optical Character recognition. OCR technology trains and enables machines to be able to extract visual data from the pictures of printed text and convert the images of text into information that can be understood by a computer program. There are two steps involved in the OCR process. In the first stage, pre-trained models are used to detect text within images. Then, Deep Neural Network models process those images and implement text recognition through the OCR module (Onkar, et al., 2019).

Object recognition is a computer vision technique capable of detecting, locating, and identifying objects of a certain class (humans, automobiles, etc.) within an image or camera feed (Mwiti, 2019). It uses image classification to assign a class label to an image and classify it into a category. Meanwhile, it applies image localization to locate the object in the video feed and draw a bounded box around it (Brownlee, 2019). Many applications on cell phones, cars, computers, and cameras have used object recognition to achieve their desired functions such as video surveillance, facial detection, object tracking, pedestrian detection, etc. through identifying objects and their locations within video frames. The Single Shot Detector (SSD) is one of the best object recognition models and uses a convolutional neural network's pyramidal feature for the efficient detection of objects of various sizes. The tasks of object localization and classification are done in a single forward pass of the network. Additionally, the MultiBox technique is used for the bounding box regression function (Remanan, 2019).

During the past years, researchers have attempted to develop all kinds of assistive aids for the visually impaired by utilizing technologies such as A.I. object and text recognition algorithms, sonar technology, global positioning systems (GPS), etc. in their designs. The Smart Reader overcame the limitation of traditional Braille by integrating an AI text-reading system with a Raspberry Pi controller and reading out the text through a speaker (Ravil, et al., 2020). Another unique device based on Raspberry Pi utilized computer vision for identifying real-time objects through the image captured by the camera module and then provided audio narration about the detected objects through an audio receiver (Abedalrahim, et al., 2022). Researchers from the Center for Research and Advanced Studies in Mexico developed AI glasses that use artificial intelligence to recognize locations, read signs, and identify objects to help the blind navigate through their environment (Borghino, 2014). Yet another type of vision aid such as a smart walking stick was embedded with Raspberry Pi, an app, a GPS system, and an alarm component. The device can emit a sound alarm regarding the obstacles detected by the sensor, use the GPS to track the position of the visually impaired and provide help via an app in an emergency (Sahoo, et al., 2019). Another aid, a walking stick used a microcontroller that emitted and received ultrasonic pulses to capture and process environmental data. By integrating the SOS navigation system into the smart stick, the device could assist the blind by providing warnings about the approaching obstacles in the path and helping them avoid emergencies (Mohapatra, et al., 2018). These revolutionary designs in past research demonstrated the capability and potential of using advanced technology to help the visually impaired regain independence in their daily life.

The purpose of this study was to build a smart visual device that could effectively help the visually impaired read printed text, identify objects in the surroundings and keep them informed of approaching obstacles through an alarm. The prototype was built by using a Raspberry Pi microprocessor and incorporating OCR text recognition, object

Vol. 2023 (2) 40 – 47
ISSN 2688-3651 [online]

recognition, and ultrasonic technology. By utilizing both artificial intelligence algorithms and ultrasonic sensors for this project, this single pocket-sized device maximized its efficiency by providing the visually impaired three types of critical information: printed text in the environment, the type of objects detected, and the distance to objects via sound and vibration. My hypothesis is that this electronic device has the potential to help the visually impaired deal with the challenges they face each day.

2. Materials and Methods

The Raspberry Pi 4B was chosen to be the microcontroller of the device and was connected to the following peripherals: Raspberry Pi Camera V2, HC-SR04 ultrasonic distance sensor, mini vibration motor disc, and earbuds (Figure 1). The OS used was the Raspbian Buster OS which was flashed onto a 16GB micro-SD card using the Raspberry Pi Imager. After the device was powered on and booted up, the initial setup and configuration were completed to connect the Pi to Wi-Fi, enable the camera, and install any software updates. To run the object recognition algorithm within a Python script, the OpenCV library, TensorFlow Lite dependencies, and Google's sample quantized SSD Lite Mobile Net V2 object detection model (trained on the COCO dataset) were installed within a Python virtual environment. The base Python script meant to run the object recognition model over the camera's live stream was downloaded into the virtual environment by cloning a public GitHub repository (Bradway, 2022). Within the same virtual environment, the Tesseract OCR library, and packages such as Pytesseract, OpenCV-python, and NumPy were also installed (*Build live text recognition with the Raspberry Pi (OCR),* 2022).

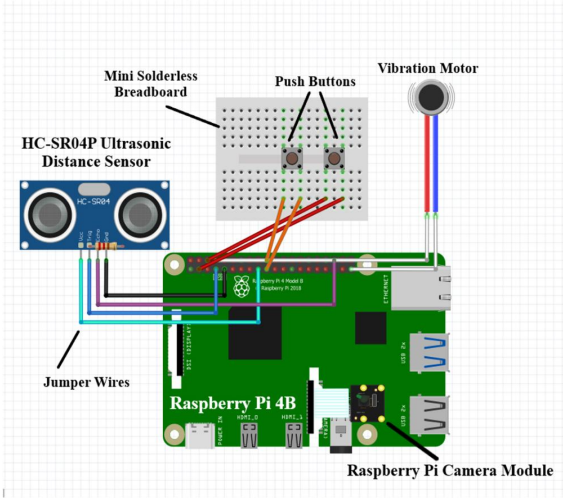

Figure 1. Image of prototype design: Pi camera, ultrasonic sensor, vibration motor, and wires were connected to the Raspberry Pi board's ports.

There were three separate Python scripts for this device. One script controlled the ultrasonic sensor which was used to calculate the distances between the device and surrounding objects. The transducer on the sensor emitted high-frequency sound waves at the speed of sound and then received them reflected from objects. The time between the emission and reception of the sound waves was multiplied by the speed of sound and then divided by two to calculate the distance between the sensors and the objects (Burnett, 2018). Depending on the magnitude of the calculated distance, the script made the mini-vibration motor discs emit vibrations at different intensities as a proximity warning. The intensity of the vibration was inversely related to the distance between the sensor and the object as the intensity increased whenever an object or obstacle was closer to the sensor.

Another Python script within the cloned GitHub repository took individual frames from the camera's live video view and processed them using TensorFlow's object detection model to recognize objects within the image. The script was modified to calculate the objects' center coordinates, which were then used to determine the object's relative position within the image frame: left, right, or forward. The code was also modified to include a text-to-speech synthesizer function which used an installed TTS engine to vocalize outputs about the recognized object type through audio output that can be heard from earbuds.

The third Python script ran text recognition using the Tesseract OCR library and Pytesseract. There were four steps in the text recognition process. In the first step, the paper with printed text messages was placed in front of the Pi camera so that the camera could capture the images of the text. The second stage of pre-processing checked whether the captured images got skewed towards the left or right, and then used the appropriate skew correction for the images. During the third step, the OCR algorithm code processed the images containing characters and identified the words.

Vol. 2023 (2) 40 – 47
ISSN 2688-3651 [online]

Lastly, the Text-to-Speech (TTS) code translated the text output into corresponding speech output through the earbuds attached to the Raspberry Pi board.

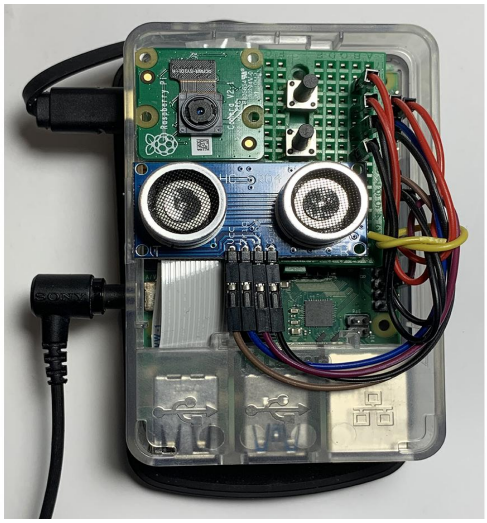

Figure 2. Image of actual prototype design: Pi camera, ultrasonic sensor, and wires were connected to the Raspberry Pi 4B board.

3. Results

The prototype design consisted of Raspberry Pi 4B board, Raspberry Pi camera, ultrasonic sensor, vibration motor, earbuds, and battery as shown in Figure 2. The Pi 4B microprocessor took the images of printed text or objects caught by the Pi camera as inputs and then produced the speech feedback of the texts or objects through earbuds as output. Additionally, the ultrasonic sensor implemented the object detection function and then triggered the motor to vibrate when the object was approaching within a certain distance.

When the Raspberry Pi was turned on, the booting screens appeared in the terminal. There are two buttons on the device. When button one was pressed, the device performed the text recognition function. While the Pi camera caught the text images and bound the images in the box as shown in Figure 3, the system programmed with the OCR text recognition algorithm processed the images, recognized the texts, and then generated speech output of the text messages that can be heard through the earbuds.

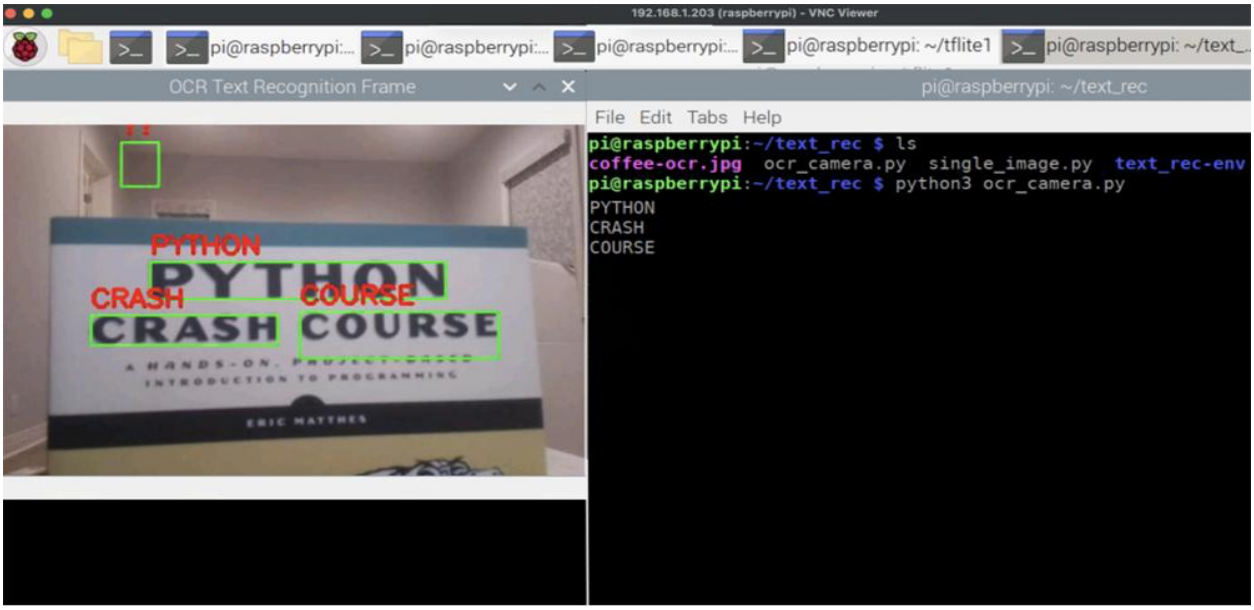

Figure 3: Image of text recognition testing: OCR algorithms recognized the words in the camera frame and indicated the text output in the terminal.

Testing was conducted by placing the paper with printed words and sentences in front of the Pi camera. The data collected from testing OCR text recognition is shown in Table 1. The collected data indicated that the device recognized all single words ranging from five-letter words to thirteen-letter words and all short sentences with 100% accuracy (Table 1). The average time taken to recognize single words was 2.5 seconds, while the average time taken to recognize short sentences was 6.4 seconds.

Vol. 2023 (2) 40 – 47
ISSN 2688-3651 [online]

Table 1: Subset of Testing Data for Texas Recognition

Text (specific words)	Audio Output	Trial 1 time (s) to successfully recognize text	Trial 2 time (s) to successfully recognize text	Trial 3 time (s) to successfully recognize text	Average Time (s)
egg	egg	2.9	3.2	3.2	3.1
cash	cash	2.8	3.2	3.1	3.03
orange	orange	2.7	2.6	3	2.77
problem	problem	2.5	2.4	2.7	2.53
assistive	assistive	2.1	2.3	2.5	2.3
technology	technology	2	1.9	2.3	2.07
competitor	competitor	1.8	2.2	2.1	2.03
environment	environment	2.4	2.3	1.9	2.2
functionality	functionality	3.1	2.4	2.4	2.63
I ate an apple.	I ate an apple.	5.1	5.3	5.6	5.33
I love to watch movies.	I love to watch movies.	6.5	6.7	7.9	7.03
We went to the beach.	We went to the beach.	6.7	6.3	7.2	6.73

When button two was pressed, the device performed the object recognition function. The window was the live feed from the Pi Camera with the object detection API. It detected the book and orange in the camera's live feed by bounding the object within a box that contained the object label and generated the text output "book" and "orange" in the window as shown in Figure 4. Meanwhile, the object location identification code produced the output of the object's relative location such as forward, left side, and right side (Figure 5). Then, the device used Text-to-Speech synthesis code to generate speech output about the object's class label and location relative to the users through the earbuds' audio output. Additionally, the ultrasonic sensors on the device calculated the distance between the objects and the device (Figure 5) and then generated vibration from motors as an alarm when the object was closer to the user within a certain distance.

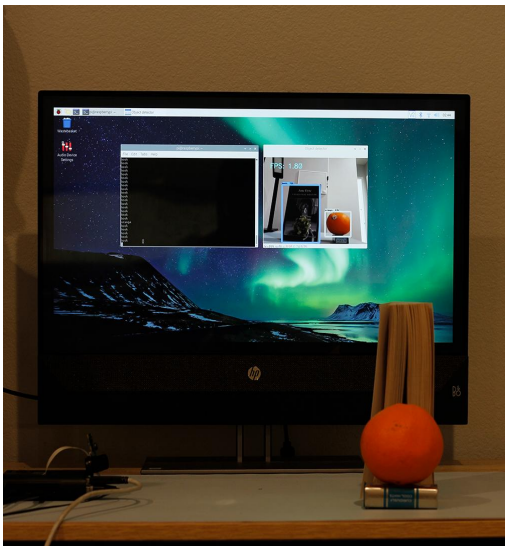

Figure 4. Image of object recognition testing: object detection model recognized the book and orange in the camera frame and printed the class name/label in the terminal.

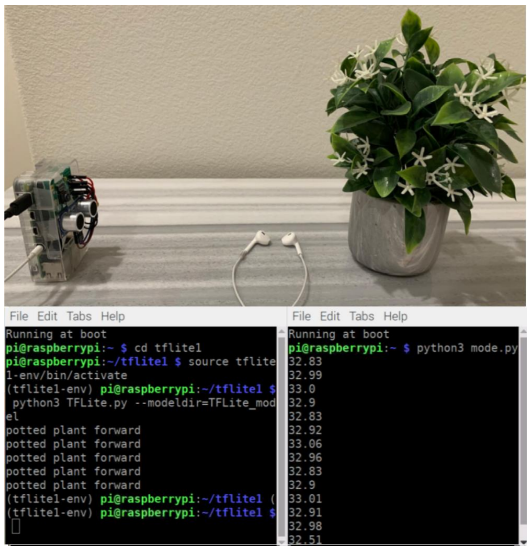

Figure 5. Images of object recognition testing: object detection algorithms identified the relative position of the objects (left, right, or forward) while ultrasonic sensors calculated the distance between the object and the sensor.

The testing data from experiments on object recognition and ultrasonic distance measurement is shown in Tables 2 & 3. The device could recognize 80% of real objects in testing correctly. It took 2.86 seconds on average to recognize

Vol. 2023 (2) 40 – 47
ISSN 2688-3651 [online]

the objects and produce speech output through earbuds. Additionally, the average maximum distance the ultrasonic sensor was able to detect the objects was 165 cm. The device could detect large objects such as cars as far as 548cm away.

Table 2: Subset of Testing Data for Object Recognition

Tested Objects	Audio Output	Actual location	Detected location	Processing time	Max. Distance Detected
Water Bottle	Bottle	left	left	4.3 s	110 cm
Laptop	Laptop	forward	forward	2.0 s	134 cm
Backpack	Backpack	forward	forward	2.4 s	123 cm
TV	TV	forward	forward	1.4 s	147 cm
Cellphone	Cellphone	left	left	1.8 s	82cm
Person	Person	left	left	1.3 s	177 cm
Fire Hydrant	Fire Hydrant	left	left	5.3 s	230 cm
Trash Bin	Trash Bin	forward	forward	3.8 s	503 cm
TV	TV	right	right	1.4 s	147cm
Person	Person	left	left	1.3 s	177cm
Streetlamp	Traffic Light	right	right	9.0 s	191 cm
Chair	Chair	right	right	2.2 s	247 cm
Plant	Potted plant	right	right	2.8 s	87cm
Car	Car	forward	forward	2.5 s	548 cm
Tree	Not Recognized	N/A	N/A	N/A	139cm

4. Discussion

The results from my experiments show that the device can quickly provide feedback to visually impaired users without sacrificing accuracy. It can read out all words in the sentences with 100% accuracy. While nearly all the objects were detected by the ultrasonic sensor, only 80% of objects were correctly identified by the device. The object recognition model had its limitations when recognizing real-life objects: it either misclassified the objects or could not recognize the objects at all. Future improvements can be made through further training of the object recognition model with more images of real-world objects. Additionally, the device could not identify the traffic light because it was not able to recognize the color patterns. The model's recognition of color and symbols such as? #, and numbers will need further research and training in order to perform these functions reliably enough for incorporation into the device.

Table 3: Statistics of Average Processing Time and Maximum Distance Detected by Ultrasonic Senso

Descriptive Statistics			
Average Processing Time (s)		Average Max. Distance Detected(cm)	
Mean	2.86	Mean	164.75
Standard Error	0.262818729	Standard Error	26.44712047
Median	2.6	Median	128.5
Mode	2.7	Mode	98
Standard Deviation	1.175361087	Standard Deviation	129.5639006
Range	4.2	Range	512
Minimum	1.3	Minimum	36
Maximum	5.5	Maximum	548

Potential future experiments could include the redesign of the device with advanced hardware which could improve the processing speed, accuracy of identification results, and addition of more functions. The current design was based on the Raspberry Pi 4B microcontroller. Replacing the Raspberry Pi board with Navida's Jetson Nano will speed up image processing and allow several computer vision packages to run simultaneously. While the ultrasonic sensor is a cost-effective solution for detecting the distance between obstacles and the user, it cannot detect the objects which are beside or beneath the sensor. On the contrary, the Lidar sensor or 3D depth sensor is compatible with the Jetson Nano microcontroller board and allows devices to provide a high-resolution three-dimensional view of their surroundings. The second version of the visual device which consists of Jetson Nano, a USB camera, and a Lidar or 3D depth sensor will enable the device to run object, text, and color recognition simultaneously without sacrificing its processing speed. To address the challenge that earbuds pose by potentially preventing the user from hearing environmental noises and sounds, a new design could incorporate both earbuds and speakers. The users can switch between earbuds and speakers according to real-life situations and the

Vol. 2023 (2) 40 – 47
ISSN 2688-3651 [online]

surrounding environment. Additionally, the user can choose to wear only one earbud, allowing their free ear to pick up the background and surrounding noise easier.

Currently, there are several vision aid devices on the market to help the visually impaired. The research on this existing alternative indicates that the AI-powered device offers an innovative solution for vision impairment. Its concept and technology-oriented design make it more effective, convenient, and accessible to visually impaired individuals than any other existing devices.

Sunu Band is a sonar wristband used in combination with a white cane to detect objects no more than six feet away. By using ultrasonic sensors to calculate the distance to objects, Sunu Band can inform users about approaching obstacles and provide prompt warnings. According to their website, the Sunu Band also pairs with an app to provide GPS assistance as real-time navigation support. However, Sunu Band cannot identify the type of obstacles in the visually impaired user's path. By utilizing TensorFlow's object recognition model that processes live video feed from the camera, the developed device not only alerts users of oncoming objects through vibration but also identifies and declares their object class via audio feedback.

Another product in the market is IrisVision's headset which utilizes mobile virtual reality by pairing Samsung Galaxy's smartphone with the Gear VR headset to help low-vision people observe objects, read and shop. The headset uses a magnification bubble to achieve magnification power so that people with low vision can see the enlarged images through goggles. However, this product is designed to help low-vision people only and is not recommended for outdoor walking because of its limited peripheral vision and depth perception. The developed device uses A.I. text and object recognition algorithms to process real-time video images from the camera and then immediately generates audio output by using a built-in text-to-speech function. It was specifically designed to target the blind and severely visually impaired, but it can also help low-vision individuals.

The Brainport Vision Pro is another alternative in the assistive aid market. It is a unique intraoral headset for the visually impaired and consists of a video camera and tongue array working with user control. It transforms the images caught on wearable cameras into electrical stimulation patterns which can be felt by the tongue. The users need to be trained to learn how to interpret the moving bubble-like pattern on the tongue to 'see' the shape and size of the objects. It is neither convenient nor easy to work with since its users must undergo complicated training programs. On the contrary, the developed device utilizes the most advanced A.I. technology thus far and is much more intuitive to operate without prior training.

Through improvements and further development, this vision aid can become a major player in the assistive aid industry and potentially assist a large number of visually impaired individuals in managing their daily challenges and regaining their independence. This study has important implications for the design of assistive aids for the visually impaired by demonstrating possible functions and indicating areas of future improvement for scientists and engineers to develop a truly reliable aid in the near future.

5. Conclusion

The development of AI text and object recognition models have enabled the construction of electronic vision aids for the visually impaired. The test results supported my hypothesis that this AI-powered device can provide a viable solution to the problem I intended to solve for this study. The device took 2.5 seconds on average to recognize single words and 6.4 seconds to recognize short sentences with 100% accuracy. This proved its potential as a text reader for the visually impaired. Meanwhile, the prototype provided quick and accurate real-time aid by verbalizing objects' names and their distance to the device with an 80% accuracy rate within 2.86 seconds on average. This demonstrated its advanced functionality and projected performance in identifying real-life objects in the surroundings. Since the device was programmed with a text-to-speech function, the text outputs from text and object recognition were automatically transformed into audio output for users to hear. The device's accuracy, effectiveness as a vision aid, and real-time speech feedback show that the prototype has fulfilled the requirement of its design criteria. Through improvements and further research, this vision aid can become a viable solution that is more effective than traditional aids such as Braille, walking canes, and guide dogs, and more affordable than existing assistive aids in the market.

Vol. 2023 (2) 40 – 47
ISSN 2688-3651 [online]

The device's innovative integration of computer vision algorithms into one pocket-sized microcontroller will enable it to address a long-overlooked problem.

References

Abedalrahim J., et al. (2022). Intelligence interfaces for assisting blind people using object recognition methods. *International Journal of Computer Science and Applications*. Vol 13 (5). https://doi.org/10.14569/IJACSA.2022.0130584

Borghino, D. (2014). *Hi-tech glasses aim to assist the blind with directions and obstacle detection.* Newatlas. https:// newatlas.com/stereoscopic-ultrasound-gps-ai-glasses-blind-assistance/32166.

Bourne, R., et al. (2017). Magnitude, temporal trends, and projections of the global prevalence of blindness and distance and near vision impairment: a systematic review and meta-analysis. *Lancet.* Vol. 5 (9). https://doi.org/ 10.1016/S2214-109X (17)30293-0

Bradway, D*. (2022) Tutorial to set up TensorFlow Object Detection API on the Raspberry Pi.* Github. https:// github.com/EdjeElectronics/TensorFlow-Object-Detection-on-the-Raspberry-Pi

Brownlee, J. (2019). *A gentle introduction to object recognition with deep learning*. Machine Learning Mastery. https:// machinelearningmastery.com/object-recognition-with-deep-learning.

Build live text recognition with the Raspberry Pi (OCR) (2022). Tutorials for Raspberry Pi. https://tutorials-raspberrypi.com/raspberry-pi-text-recognition-ocr.

Burnett, R. (2018). *Understanding how ultrasonic sensors work.* MaxBotix. https: //maxbotix.com/articles/how-ultrasonic-sensors-work.htm

Mohapatra S., et al. (2018). *Proceedings of 2018 2nd international conference on trends in electronics and informatics*. IEEE Xplore. https:// ieeexplore.ieee.org/document/8553935.

Mwiti, D. (2019). *A 2019 guide to object detection.* Heartbeat. https:// heartbeat.fritz.ai/a-2019-guide-to-object-detection-9509987954c3

National Center for Health Research (2004). *Blind adults in America: their lives and challenges.* https:// center4research.org/blind-adults-america-lives-challenges.

Onkar, O.S., et al. (2019). Optical character recognition for the blind using Raspberry Pi. *International Research Journal of Engineering and Technology*. Vol. 6(3). https://www.irjet.net/archives/V6/i3/IRJET-V6I3520.pdf

Ravil, A., et al. (2020). Raspberry pi based smart reader for blind people. *Journal of Analysis and Computation (JAC)*. Vol. XIII (III). https://ijaconline.com/wp-content/uploads/2020/06/1012-Paper-Dr.-A.-Ravi.pdf.

Rein, D.B., et al. (2022). The economic burden of vision loss and blindness in the United States. *Ophthalmology.* Vol. 149 (4). 369-378. https://doi.org//10.1016/j.ophtha.2021.09.010

Remanan, S. (2019). *Beginner's guide to object detection algorithm.* Medium. https://towardsdatascience.com/beginners-guide-to-object-detection-algorithms-6620fb31c375

Sahoo N., et al. (2019) Design and implementation of a walking stick aid for visually challenged people. *Sensors*. Vol. 19 (1). https:// doi.org/10.3390/s19010130 mdpi.com/1424-8220/19/1/130.

The cost of a service dog. (2022). Guide Dogs of America. https://www.guidedogsofamerica.org/cost-of-a-service-dog/

Voulodimos, A., et al. (2018). Deep learning for computer vision: a brief review. *Computational Intelligence and Neuroscience*. Vol 2018. https://doi.org//10.1155/2018/7068349.